U. S. NAVAL SCHOOL OF AVIATION MEDICINE
NAVAL AIR STATION
PENSACOLA, FLORIDA

JOINT PROJECT REPORT NO. 19

The Ohio State University Research Foundation
Columbus, Ohio, under Contract N6ONR 22525,
Office of Naval Research Project Designation No. NR 145-993

and

U. S. Naval School of Aviation Medicine
The Bureau of Medicine and Surgery Project NM 001 064.01.19

APPLICATIONS OF MULTIPLE-CHOICE SPEECH INTELLIGIBILITY TESTS IN THE EVALUATION AND USE OF VOICE COMMUNICATION EQUIPMENT

Report prepared by

John W. Black, Gilbert C. Tolhurst,

and

S. N. Morrill

Approved by

John W. Black
Project Director
and
Captain Ashton Graybiel, MC, USN
Director of Research
U. S. Naval School of Aviation Medicine

Released by

Captain James L. Holland, MC, USN
Commanding Officer

15 November 1953

Opinions or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department. Reference may be made to this report in the same way as to published articles noting authors, title, source, date, project number and report number.

SUMMARY

Multiple-choice intelligibility tests were developed during World War II for measuring the results of training in voice communication. The tests have been applied only in situations in which only an occasional measurement was required that employed the same personnel, as before and after classroom instruction. This restriction arose because of the inflexibility of the multiple-choice tests and the restrictions imposed by a single formal printed answer form. More recently, additional forms have been devised and the tests have been used extensively in evaluating equipment and the performance of operators. Summaries of approximately 20 such applications comprise this report. The topics of typical uses include evaluations of visual monitoring of voice level, temporal regularity of voice signals, the relative intelligibility of recorded speech, a relationship between side-tone and "free-room" intelligibility, the components of the tower-to-aircraft radio system, comparisons of alternative headsets and microphones, and bone-conduction masking.

A singular utilization of the multiple-choice tests is through controlling the separation time of the test items and determining the rapidity with which the test items can be identified through comparison systems.

INTRODUCTION

The several components of a voice communication system commonly require evaluation with respect to their efficiency in transmitting speech. The multiple-choice word intelligibility test provides a measure of the relative proficiency of both equipment and speaker-listeners, as in a comparison of scores earned before and after training. One set of these tests was devised during World War II (11) and a similar set has been developed more recently at this Laboratory (4). Each set consists of a pair of 12 lists of either 24 or 27 word items with the lists equated for mean difficulty and variance of the relative reception scores of the individual items. A listener responds to three words that are read with an introductory or carrier phrase, the five words comprising a single unit. His response to a test item is to draw a line through the word he hears, one of a group of four "option" words that appear together on his answer form. These four words include the test item and the three words that were written most frequently as mistakes when the stimulus word was presented in write-down tests.

Historically, intelligibility tests were employed in the evaluation of communication equipment before they were used in measuring the speaking efficiency of talkers. Engineers of the Bell Telephone Laboratories tested the components of telephone systems in this manner, and expressed results relative to an orthotelephonic system, i.e., the articulation score in quiet at 100 cm. through air in a sound-treated room (8). (An electrical equivalent of the orthotelephonic system was constructed.) Speech tests were widely used during World War II to evaluate alternative equipments. The typical tests were write-down as to methods of

responses using word, syllable, or sentence types. The inter-relationships among these three were formulated (6,7).

One purpose of the present report is to present through illustrations the usefulness of the multiple-choice intelligibility tests in measuring and appraising the voice signal and communication equipment. No advantage is claimed for these tests over conventional write-down forms with regard to either validity or reliability. The multiple-choice tests, however, are more economical in time of scoring and seem to offer a more agreeable task than write-down tests to experimental subjects. This latter fact may be of consequence only in circumstances that involve amateur experimental subjects. Either a template or IBM equipment may be used in scoring listeners' responses, depending on the type of answer form that is employed. A necessary precaution in interpreting the examples of the present report relates to the "circular" implication that since the test distinguishes among experimental conditions (a) the test is "good" and (b) the conditions are "dissimilar." This precaution is especially relevant since most of the illustrations herein are presented as experimental findings--a second purpose of the report--as well as examples of the usefulness of the multiple-choice tests.

The applications of the multiple-choice tests that are discussed with respect to the evaluation of equipment and its operation are grouped under four headings: (a) "voice signal," a comparison of two or more modes of performance by the operator; (b) direct comparisons of alternative equipments; (c) "response readiness," a comparison of the scores of listeners when they hear especially recorded tests, controlled in speed, via alternative equipments; (d) "efficiency in noise," a comparison of the scores of listeners when hearing recordings of two tests via alternative equipments as reproduced through an artificial voice in a controlled noise environment.

The majority of researchers who work with problems involving intelligibility are accustomed to situations that involve a limited number of trained listening crews. Some of these crews work professionally as listeners. The studies of the present report were conducted with panels of listeners who served as experimental subjects at most for three or four different experiments following a brief period of indoctrination or training.

Four forms of the multiple-choice tests are referred to as Forms A, B, C and D. Forms A and B are the Haagen test, 24 items per list, and Forms C and D are the ones locally developed, 27 items per list (11,4).

APPLICATION 1: VOICE SIGNAL

Summaries of tests related to speaker performance follow. In these instances the speaker is viewed as an operator of voice communication equipment.

A. Visually monitored voice level. Ninety-six male experimental subjects--research personnel, unfamiliar with the tests except for

indoctrination and practice in the testing procedure--read the intelligibility tests of Forms A and B. Each subject read from a small sound-treated room in the presence of 85 db of recorded aircraft noise. He spoke into an Altec-Lansing 21C chest-plate microphone, positioned six inches from his mouth. (Directions for the speakers are reproduced in Appendix A.) The side-tone under the earphone cushions of the reader was approximately 105 db, the peak value indicated from a probe tube-condenser microphone system and read on a Hewlett Packard voltmeter (Modd 400-C).

Panels of listeners sat in tablet arm chairs in a sound-treated room apart from the speakers and in the presence of 110 db of simulated aircraft noise, propeller type.* The listeners heard the tests through HS-33 headsets. The modal peak value of the speech-to-noise ratio under the listeners' earphones was 10-12 db as indicated by the probe tube system.

The object was to test the effect upon intelligibility of a visual monitoring system for the signal level of the speaker. Commercial radios and voice recorders typically include a visual monitoring device that serves both to warn the operator about a signal level that is too weak to be readable and to protect against overdriving the equipment--in other words, to increase the likelihood of control of the level of the signal at the source. The present comparison was to explore the possibility that listeners' speech-reception scores might be affected by unmonitored and varying voice levels in the same manner as equipment, and that a visual monitoring device might serve to stabalize the sound pressure level of the speaker.

One half of the experimental subjects read under the standard directions (Appendix A). The remaining half read with the additional instruction, "Keep the needle of the VU meter at minus 2." The prescribed relative level was the typical value of the signal when a preliminary group of subjects read the tests under standard instructions. Hence, it was not necessarily the optimum level. The order of these experimental conditions was reversed from one panel of 12 speaker-listeners to another.

---

* The rooms that were used in the studies of this report have a noise level in quiet of 27 db (General Radio meter, A scale) and an isolation attenuation of 55 db. The walls and ceiling are faced with Johns-Manville "Transite." The floors are carpeted.

The noise of propeller-type airplanes is produced by a Harvard generator; white noise originates with an H. H. Scott noise generator, Type 810-A.

A 50-watt amplifier feeds four Altec-Lansing 287-W amplifiers, 250 watts each. Each 287-W amplifier drives one of a bank of four Altec-Lansing 515 low frequency reproducer units mounted in a modified A-2 baffle with four Altec 290-B high frequency drivers mounted on top of the 515 horns dispersed through a multi-celled horn and four Altec 802-C "super-tweeters" mounted within the flair of the A-2 horn.

Results. There was no statistically significant difference between the intelligibility scores that accompanied the two conditions : means, "monitored" 68.9% vs. "not monitored" 67.7%; t = .50 (94 d.f.). There are various possible explanations of this negative result including the one that the test was not sufficiently sensitive to detect a difference. This latter possibility seems improbable in view of B (below). Other explanations would be that the acoustic monitoring system supplies cues that lead to both an adequate sound pressure level and an even level; and that a visual system is "not followed anyway." As stated above, the selected reference level was not necessarily the optimum one. Therefore, the view that the acoustic monitoring system yields an "adequate level"lay outside the scope of the study. The possibility that the selected level tends to be maintained evenly is consistent with B (below) and the indication therein that the speaker is sensitive to the acoustic feedback. The question of whether the visual system is followed or not leads to an examination of the standard deviations of the arrays of scores. Presumably an effective visual system would increase the level of "weak" voices and decrease the level of "strong" voices and, in turn, reduce the variability of intelligibility scores. The obtained standard deviations were: visually monitored, 9.5%; not visually monitored, 11.8%. These values might indicate that the speakers followed the system and that variability among voices was reduced without affecting the mean intelligibility of the group.

B. Side-tone: acoustic reflection to the ear in quiet. The sound pressure level of the speech signal is a major determiner of intelligibility (1,3,11,12,13). The level at which a speaker talks is, in turn, affected by conditions that determine the level with which a speaker hears himself, in other words his side-tone or acoustic feedback. Rooms of differing sizes and reverberation time are known to alter the level at which a person talks (2). Whether or not this change in level is sufficient to affect the speaker's intelligibility has not been determined. Moreover, the multiple-choice intelligibility tests were constructed for use with a noise barrier that would attenuate the scores to the mid-range of the scale. The level of this barrier was normally 110-114 db in an "intercom" situation and 68 db in a "free-room" circumstance.

An experimental condition was desirable that would (a) accommodate a listening panel and (b) provide a "minimal" change in the speaker's acoustical environment--somewhat comparable to a change induced by "acoustical treatment" vs. "no acoustical treatment" in a room. With these conditions present the effect of altered circumstances of side-tone upon "free-room" intelligibility might be tested.

A "quiet" corridor 86' x 7' x 8' was made available under an arrangement in which only the experimental subjects and the experimenter were in the building. The ceiling of the corridor was sound treated, the walls of hard plaster, and the floor of polished hard wood. Fifteen tablet arm chairs were placed at right angles to one wall of the corridor, spaced four feet from center to center, with one chair 26 feet from the "speaker end" of the hallway. At this end two facing doors led to rooms that were adjacent to the hall. One of these doors was of hard surfaced, polished wood; the other was covered with felt and opened into a sound-treated room (5 surfaces, acoustic tile; the floor, carpeted).

Twelve panels of 15 experimental subjects participated in a series of testing situations in which 12 members of each panel read the intelligibility lists of Forms A and B. Each speaker read two times from a position midway between the two doors, four feet from the end of the hall. The felt-covered door to the sound-treated room was always open. In one instance each speaker read "naturally" in this air line. In the other experimental condition the subject read with an Ear Warden (V-51R) fitted in one ear. For half the subjects the Ear Warden was in the ear facing the polished door; for the other half, in the ear facing the sound room.

The order of conditions (Ear Wardens vs. no Ear Wardens) was counterbalanced among panels and the order of "Ear Warden in the right ear" vs. "Ear Warden in the left ear" was counterbalanced among the "Ear-Warden" conditions. Forms A and B of the intelligibility tests were counterbalanced among the several conditions of reading.

An assumption in this investigation was that the speaker would adjust his speaking level, in part, by the sound pressure level of the sound waves that were reflected from the polished door and impinged on his ear. Thus, except as factors such as differences in the level of bone-conducted side-tone might have to be taken into account, an Ear Warden in the ear away from the door would not affect speaker intelligibility (through an alteration of vocal sound pressure level) while an Ear Warden in the ear facing that door would effect intelligibility.

Results. All mean speaker intelligibility scores for the different conditions were high, ranging upward from 87%. Such scores were to be expected in view of the absence of a noise barrier. The mean scores for 144 speakers in the "no Ear Warden" vs. "Ear Warden" conditions were 87.4 and 88.9% respectively (t = 2.80, significant at the 1% level of confidence). The mean scores for the 72 subjects in the "no Ear Warden" vs. "Ear Warden facing the open door" conditions were 87.7 and 88.6% respectively (t = 1.14, not significant). The mean scores for the 72 subjects in the "no Ear Warden" vs. "Ear Warden facing the 'reflecting' door" were 87.4 and 88.9% respectively (t = 2.57, significant at the 1% level of confidence).

These results would suggest that (a) the monitoring signal for the speaker came largely from the reflecting surface, approximately three feet from his right ear, and that (b) when this surface-ear exterior pathway was rendered ineffective, the speaker, employing some other cue for "natural level," increased his voice level sufficiently to affect his intelligibility score to a statistically significant degree. No measure of this change in physical units was available. The difference in intelligibility scores, however, was equivalent to the difference in "listener-separation" distance of 12 feet in the row of seats from 22 to 32 feet from the speaker.

Modification I. The study described immediately above was conducted in quiet under an acoustic condition of 55 db, achieved after turning off all motors in the building, including an air conditioning unit and a water cooler. The experimental procedure was repeated

with the building noises present, but again with only experimental personnel in the building. The mean of the speaker scores, without Ear Wardens, N 144, was attenuated from 87.40 to 86.15 (t, 1.65; significant at the 10% level of confidence). Moreover, no significant change occurred in the intelligibility scores when an Ear Warden was inserted in either ear of the speaker. This might indicate that the auditory masking of the building noises was sufficient to negate the effect of the reflecting door as a principal source for the monitored signal.

Modification II. The experimental plan outlined above was followed a third time with portable acoustic reflectors introduced in the transmission line of the airborne side-tone. A plywood box 48" x 30" x 10" was "cut" to form two circular faces (48" radius or an arc 9" deep), one section concave and the other convex. The "back" faces of the sections were flat. A flat side of one of the sections functionally replaced the reflecting door of the studies above and was arbitrarily placed 12 inches "from the speaker's ear." Alternately the convex and concave surfaces replaced the open door (sound room) of the studies above. On the inter-aural axis these surfaces were also 12 inches from the speaker's "other ear." The assumption under test was that the concave reflector would tend to focus sound waves at the speaker's ear, cause him to attenuate his voice level and consequently reduce his intelligibility score. The convex surface, on the other hand, would tend to disperse sound waves and cause the speaker to "talk up" (or alternatively transfer the source of the monitored signal to the ear facing the flat surface). In either instance the vocal signal would be increased in level and consequently in intelligibility.

Results. The mean intelligibility score of the 72 speakers who read with a convex or sound-dispersing unit at their left ear was 90.6, and of the 72 speakers who read with a concave or sound focusing unit, 89.4 (t, 3.03, significant below the 1% level of confidence). Presumably the concave surface reflected sound to the ear and provided the source for the side-tone that was monitored with the result that vocal level was attenuated. However, when the convex surface faced the left ear, distributing the sound, the flat surface at the right ear may have become the source for the monitored external side-tone. In any event, the relatively less efficient acoustic channel from the mouth to the ear led to an increased intelligibility score, presumably due to a greater output level from the mouth.

C. Level of side-tone (13). The "intercom" system of the naval aircraft that is used in the first stage of flight instruction feeds the same level of signal to the headsets of the talker and to the listeners on the "party line." Commercial telephone systems are constructed so that they attenuate the level of the speaker's side-tone. The possibility arises that a similar reduction in the side-tone of the military headset in aircraft might improve intelligibility through altering the level of speech and thus improve the speech-to-noise ratio at a listener's headset.

Each of 16 experimental subjects read four lists of Forms A and B of the multiple-choice intelligibility test to panels of 12 listeners. Both listeners and speakers were surrounded by 110 db of simulated aircraft noise, propeller type. Each speaker experienced four levels of side-tone as he read: "normal" (approximately 105 db as described in A, above), - 14, - 27, and - 38 db. The order in which the four levels were presented was counterbalanced among the 16 speakers. In addition to intelligibility scores, a relative measure of vocal level was obtained from graphic recordings of the spoken materials (Sound Apparatus Co. power level recorder).

Results. Both the mean sound pressure level of the speakers and intelligibility increased systematically with decreased level of a side-tone. An increase of 13 db in vocal level was occasioned by the 38 db attenuation in side-tone, an overall ratio of 1:3 db. In successive steps the ratios (vocal level increase: side-tone attenuation) that resulted from the attenuation settings were 1:2, 1:3, and 1:4 db. Thus the first decrement from "normal" side-tone produced the greatest increment in voice level. The increment in vocal level that accompanied the reduction in side-tone bore an overall relation to intelligibility of 2 db = 3%. In successive steps from "normal" si -tone this was 1 db: 1%; 2 db: 3%; and 1 db: 4%. Thus, although the final step in the attenuation of side-tone (11 db) did not produce a marked increase on vocal level (2-3 db), this increment in vocal level accounted for the most marked single rise in intelligibility (8%).*

D. Regularity in the signals. The instruction "talk naturally" is commonly included in a final phase military voice-communication courses. The direction has special relevance to patterns of vocal inflections and groupings of words and is intended to apply to aspects of talking that are not specifically treated in preceding sections of the training program in voice communication. One assumption is that "irregularity" within an uttered sequence of closely related words deters intelligibility. Possibly this view derives from the common notion that irregularity is detrimental to the "prettiness" aspect of speech and possibly from the "value" that is placed on fluency. The multiple-choice intelligibility tests are comprised of three-word groups, each read with a carrier phrase ("number 1" etc.) as though the entire unit were a five-word sentence. Thus the test itself provides stimulus material through which the effect of irregular temporal spacings of words might be tested with respect to intelligibility.

Forms A and B of the intelligibility tests were recorded on magnetic tape in quiet by one trained speaker. The speaker varied the ordinary manner of reading a list by saying the three test words of a group as discrete items, i.e., with a stoppage of breath after each item. These recordings of Forms A and B were copied two times. One

---

* Subsequent to this study and with the cooperation of Dr. Harvey Fletcher, a photostatic copy of an early report on the topic of side-tone and voice level was made available to the authors (9). The report of 1918 contains quantitative results that relate directly to the ones of the present study.

set of "copies" was edited to introduce a silent interval 0-4 sec. between items one-two and between items two-three of each three-word group. The intervals were the quarter-second values within the range (0-4 sec.) and were introduced in random order.

The four recordings, two "regular in time pattern" (uncut) and two "irregular in time pattern" (cut) were played back to 96 experimental subjects, 24 hearing each recording. (The Directions for Listeners in the multiple-choice tests are given in Appendix B.) The listeners sat in 114 db of simulated aircraft noise, propeller type. The signal-to-noise ratio, peak r.m.s., under the ear-phone was 12-15 db.

Results. The results of this comparison were negative. The respective mean intelligibility scores were: irregular, 81.9%; regular, 81.0%; t, 1.14, nonsignificant. Apart from the principal comparison, regularity vs. irregularity in the spacing of the items, the results agree with other indications that no penalty to intelligibility score accrues through presenting the items in rapid succession rather than spread out over a "comfortable" period of time.

E. Recorded vs. live signals. Frequently in intelligibility testing, a recorder is placed in the communication system with a view toward reproducing the voice signals as experimental stimuli later. This arrangement is essential in the automatic programming of stimuli. Also the stored stimuli may be subjected to acoustic analysis. The effect of the recorder-reproducer link on speech reception scores was tested.

Seventy-two male subjects in groups of 12 read either Form A or B of the multiple-choice tests. The reading occurred over a mock-up of an aircraft intercommunication system in the presence of 110 db of simulated aircraft noise, propeller type. As a panel read, the lists were recorded simultaneously on both a Presto 24L disc recorder and a Stancil Hoffman R-4 tape recorder. The speakers were heard directly over the "intercom" system by their fellows who served routinely as a listening panel. (This manner of administering the tests is referred to as round robin.) The output of the amplidier of the "intercom" system was recorded. The recording engineer monitored the signal level of the headset circuit with a VU meter. A panel of listener-speakers, having heard each other read Form A (or B) of the test, then heard the alternate test (Form B or A) as it had been recorded by the preceding panel. The original level at the listeners' headsets was maintained. This continued through six panels. Subsequently, other panels of listeners heard the remaining set of recordings.

Results. The three sets of speaker scores were treated by analysis of variance. The F-ratio for conditions ("direct" scores and scores from two recorders) was highly significant, F = 29.0, (2 and 71 degrees of freedom).* The respective mean intelligibility scores for the three

---

* The basic measure in all analyses of this type in this report was an average score that was derived from pooling the responses of the listeners with respect to one speaker.

conditions were: direct, 68.1%; disc recorder, 64.5%, and tape recorder, 64.4%. Both "recorder scores" were significantly lower than the "live" scores, but were not significantly different from each other.

The scores yielded by the three systems were then tested to find whether the speakers' scores were only attenuated by the recording process or altered in relative merit as well. Product-moment correlations of the 72 speaker scores were: live <u>vs</u>. disc recording, <u>r</u>, .83; live <u>vs</u>. tape recording, <u>r</u>, .81; disc <u>vs</u>. tape recording, <u>r</u>, .88. Correlation values of this magnitude are common in split-half and test-retest correlations and under the conditions of the experiment indicate that the recording-reproducing process may be introduced into intelligibility testing with caution.

<u>F. The representative voice</u>. Common practice in intelligibility testing emphasizes the generalized ear. This ear is provided by multiple listeners, a crew or panel. Frequently, however, the same emphasis is not placed on the generalized voice. Indeed, in the present report instances are reported in which one voice was utilized. The risk involved in this practice is implied in the tests of <u>Application 4</u> (below) in which instance two voices recorded the speaker lists of one form of the intelligibility test. The recorded tests were reproduced to listeners in varying circumstances of listening. As the conditions of listening became more difficult, the responses to the two voices were affected differently over the comparison equipments: an interaction between equipments and voices.

The usual method for obtaining the generalized voice with the multiple-choice tests is through the use of the round-robin administration of one form of the test with each panel of 12 listener-speakers. Alternatively, either multi-speakers may read the lists, as many as 12 speakers per form, or multi-voices may read the items within each list. In one recording that was prepared in the latter manner, nine speakers were equally spaced about a nondirectional microphone and spoke with similar sound pressure levels. Each speaker read one line from each speaker list. The speakers rotated their order systematically from list to list in a manner that permitted each speaker to read from first to ninth position in nine of the 12 successive lists. This procedure built speaker variability into the recorded lists and accomplished the generalized voice within the test materials employed.

Results: The lists that were recorded in the manner described above were not different in intelligibility as "scored" by 12 panels of listeners. The nine speakers, differing significantly, were assigned scores ranging from 55% to 61% (mean 58.1 ± 2). The same listeners who responded to the "generalized voice" also heard a single trained voice read a comparable form of the test. The respective mean scores were: generalized, 58.1%; single voice, 77.3%. The procedure of building the generalized test into the successive lists appears to be a rewarding approach where applicable.

APPLICATION 2: THE INTELLIGIBILITY INDEX OF EQUIPMENT OUTPUT

The distinction between the materials that are summarized under this heading and the ones that precede and follow is largely in the emphasis of the different studies. The focus here is on equipment, not the operator, and in the intelligibility of the output of the equipment under simulated conditions of operation. Five illustrative studies are summarized.

A. Stages from the tower to the aircraft radio. One voice recorded intelligibility lists from Form B "tapped off" from various links in a tower-to-plane radio system: (a) after the microphone preamplifier, (b) after the limiter, (c) after the underground transmission line from the tower to the transmitter, (d) after the modulator section of the transmitter, and (e) from the aircraft receiver. The recording was made under the "natural" ambient noise conditions associated with the station at which the recording occurred. The recorded tests became parts of a demonstration recording with descriptive expository materials interspersed and with extra lists utilized for comparisons of alternative equipments at the stations. The recording served as a convenient indoctrination exercise in explanations of equipment, as the speech signals could be "administered" in the manner of a routine intelligibility test. The results of these administrations in quiet were retained for analysis. Subsequently, the test lists were edited from the recording and administered to 39 listeners in the presence of 114 db of simulated aircraft noise, propeller type, and again to 37 listeners with white noise introduced into the headset line (-4 db signal-to-noise ratio, peak r.m.s. values).

The alternative equipments that were employed included: (a) a condenser and a carbon microphone, (b) two pairs of transmitter-receiver combinations, an MHF-TDD transmitter (382 kc.) operating with both 40% and 90% modulation and an associated receiver, and a VHF-TDQ transmitter (125.5 mc.) and an associated ARC 1 receiver.

Results. The recordings of the stimuli of this study were necessarily made in different field locations; and were made with single samples of each piece of equipment, and the preceding intermediate equipments. The principal controls were (a) the same voice throughout and (b) a voltmeter with which to monitor the input of the recorder. The recorded stimuli were copied on a power level recorder (Audio Devices, Logger). This showed that the recordings "in the tower" were of the same level ±1 db; the recordings "outside the tower" varied ±6 db and were less intense (median difference, 6 db) than the ones "in the tower." Hence, except for a preliminary analysis, the two sets of data were treated separately.

An analysis of variance was performed with the mean reception score of a listening group to one condition in quiet constituting a basic measure. Thus, columns represented conditions (equipments) and rows, panels. The analysis yielded highly significant differences among conditions, $F = 10.21$ (11 and 9 d.f.). The comparable reception scores for the panels "in noise" were more variable from condition to condition than the scores "in quiet."

In the tower. Five test lists were recorded in the tower: (a, b) at the output of a condenser microphone, same pre-amplifier as in c (two tests); (c) at the output of the carbon microphone pre-amplifier; and (d, e) at the output of the limiting amplifier when fed by the condenser microphone (two tests).

The following values are relative mean intelligibility scores that were obtained with the listeners in noise. The condenser microphone at the microphone stage provides a reference intelligibility score. The corresponding sound pressure levels as indicated by the graphic level recorder are also enumerated.

	Relative Intelligibility Scores	Relative SPL Values (Logger)
(a,b) After microphone stage - condenser	0%	0 db
(c) - carbon	-14%	0 db
(d,e) After limiter - condenser	-25%	+1 db

Outside the tower. Seven tests were recorded outside the tower: (a) at the output of the underground transmission line to the transmitter; (b, e) at the output of the modulator section of the HF transmitter set at 40% modulation, and--through air pickup--at the output of the associated aircraft receiver ARC-5 with the plane on the ground; (c, f) same, HF transmitter with 95% modulation; and (d, g) at the output of the modulator section of the VHF transmitter, 128.5 mc., 80% modulation, and--through air pickup--at the output of the associated aircraft receiver ARC-1 with the plane on the ground. The test list that was transmitted from the tower in all of these instances was initiated through a condenser microphone.

The following relative articulation and sound pressure level values are to be interpreted in the manner of the values in the tower (above). Thus, with the listeners in noise the VHF transmitter yielded the highest intelligibility scores and the HF transmitter, 40% modulation, the highest sound pressure level.

	Percent Relative Scores	Relative SPL Values
(a) Underground Transmission Line	- 4%	0 db
(b) Transmitter HF, 40% modulation	- 1%	- 5 db
(c) Transmitter HF, 95% modulation	- 7%	- 6 db
(d) Transmitter VHF, 80% modulation	0%	-12 db
(e) Receiver ARC 5; Transmitter 40% (b)	-26%	- 7 db
(f) Receiver ARC 5; Transmitter 95% (c)	-28%	-10 db
(g) Receiver ARC 1; Transmitter 80% (d)	- 4%	-10 db

Differences between pairs of relative intelligibility values of 5% may be accepted as highly significant statistically. Thus, within the limits of a single system, one voice, and the listeners responding in 114 db of noise some comparisons are possible when qualified by the relative signal levels of the tests. The condenser microphone yielded better scores in the tower than did the carbon microphone; however, the present "limiting stage" is not well suited to the condenser microphone. Also, increasing the percent modulation of the MHF transmitter from 40 to 95% apparently did not benefit the listener, although the differences in signal level prevent a definite comparison. The VHF transmitter and its associated ARC 1 receiver yielded higher reception scores than the HF equipments.

E. Speech reception via an insert-type headset (universal insert). Conventional military headsets are criticized as bulky, heavy, and inconvenient. More importantly, the headset is a link in the imperfect communication system of the aircraft. Novel headsets frequently "catch on." The headset that was tested and is the subject of this section had "caught on" and reportedly was being purchased by a number of pilots as personal property. The earphones were conventional hearing aid receivers, mounted on a light-weight plastic-covered steel band. The signal was fed to the ear through a metal tube 2 inches in length and 1/8 inch inside diameter that terminated in a small plastic "marble" with a hollow axis of 1/8 inch. This porcelain-like terminal fitted snugly against the end of the ear canal.

One voice recorded four lists of Form A of the multiple-choice intelligibility tests. The recording was made in quiet with a high-quality microphone. Twenty-four groups, each of four members, served as listeners. One listener wore the test headset and another the conventional headset (HS-33). The third and fourth members also wore this pair of headsets over V-51R Ear Wardens. The listeners sat in 110 db simulated aircraft noise, propeller type. The same four listening stations were used throughout and were rotated among the four experimental conditions from group to group.

Results. The reception scores of the 96 listeners were treated through an analysis of variance: headsets, 1 d.f.; ear wardens, 1 d.f.; interaction, 1 d.f.; and "within cells," 92 d.f. The respective variances were 413, 1292, 4, and 144. The F ratio was highly significant in the comparison of "Ear Warden" vs. "No Ear Warden," and was significant below the 5% level of confidence in the comparison of the two headsets. The mean values of the listening scores under the four conditions were:

Headset	Without Ear Wardens	With Ear Wardens
HS-33	85%	80%
Test Set	83%	75%

The comparison insert headset displayed no advantage in intelligibility, and under the circumstances of the experiment, ear wardens were deleterious to speech reception. This last fact apparently differs from

come previous results, although the benefits in signal reception that have been found to accrue from the use of ear defenders is a function of noise level and signal-to-noise ratio (3,12,15).

C. Speech reception via an insert-type headset (fitted inserts). Another insert-type headset, one that represented a considerable departure from conventional designs, was tested. It contained no headband, only two hearing-aid type receivers. These were fitted into two hearing-aid type ear molds, individually fitted to the wearer, and held in position by the structure of the pinna and the external auditory meatus. Ear Wardens would not be feasible with this headset, the receiver attachments forming a close seal at the ear in the same space that would be occupied by ear defenders of the V-51R type. Therefore, the subjective tests of the headset, insofar as they involved speech reception, were limited to a direct comparison between the new headset and the HS-33 headset. The test headset yielded a 10 db greater output than the HS-33 when the same signal fed the two. The level of the received signal being subject to the control of the operator in practice, the two headsets were equated prior to word-intelligibility tests. The equating was done through loudness balance tests by six experienced listeners with normal hearing. The reference test signal was a 1000 c.p.s. tone and the subjective criterion level was "maximum comfort" with the HS-33 headset.

The 12 lists of Form D of the multiple-choice intelligibility tests were recorded in quiet by three male voices. The three readers spoke the lists in rotation. Twenty-four experimental subjects heard the recorded tests. Half of the lists were heard through each of the headsets under comparison. White noise was introduced into the headset circuit in an amount to make the peak r.m.s. signal-to-noise ratio approximately zero. In addition, the listeners sat in 114 db of simulated aircraft noise, propeller type.

Results. There was no significant advantage of one headset over the other. The mean speech reception scores for the two units were: test insert headset, 25.9%; HS-33 headset, 23.0%

A further comparison was made to find whether or not the individually fitted ear inserts yielded higher intelligibility scores than the same headsets without individually fitted ear molds. Each listener responded to another form of the multiple-choice tests, using fitted ear pieces for half the lists, and not-fitted ones for the other half. In this comparison there was no in-circuit noise. The fitted ear molds yielded significantly higher scores than the ones that were not fitted. The mean speech reception scores were: fitted, 75%; not fitted, 56%.

D. The relative intelligibility of microphones. One treatment in the testing of the relative efficiency of microphones is to test them with speaker-listener panels. Unlike the headset, the microphone is "fixed" in input-output ratio insofar as the speaker-operator is concerned. Moreover, a particular microphone is to be considered as matched to its amplifying system. Hence, the test cannot employ alternative microphones simply interchanged in the same circuit. A practice that is followed in these tests is to establish microphone circuits that deliver the same voltage to the headset in quiet when a speech signal

is impressed on the microphone and to accomplish this with minimal overloading or distortion in any part of the system. A convenient reference signal level is the modal r.m.s. peak value that is delivered by the carbon microphone to the headset in the "intercom" system of the aircraft that the student pilot flies first. When the multiple-choice intelligibility tests are recorded through a high fidelity microphone and heard in quiet over HS-33 headsets, the tests yield a score of 96%. With the carbon microphone the score is approximately 92%. The reception scores are further attenuated by reducing the level of playback, by the introduction of masking noise in the circuit, or by surrounding the listener with noise.

Twelve experimental subjects read the tests of Forms A, B, and C of the multiple-choice intelligibility tests in conjunction with three microphones: the hand-held carbon microphone (RS-38), the noise cancelling carbon microphone (M-UR/6), and a commerical condenser microphone. Eleven listeners responded to the tests as they were read. The three forms of the test were equally represented among the three microphones. (This procedure is not recommended, Form C yielding lower scores than Forms A and B. This fact had not been ascertained when the present application was made.)

Results. An analysis of variance of the speaker scores yielded the following values: variance, microphones 144.34, speakers 112.30, remainder 23.08 (2, 11, and 22 d.f. respectively); F, for microphones, 6.25 (highly significant). The respective mean scores associated with the three microphones were: hand-held, 70.0%; noise cancelling, 76.2%; and condenser, 75.8%. Both of the latter two values were statistically significantly higher than the value for the hand-held microphone but not different from each other.

E. Bone-conducted masking (5). An evaluation was made of a flight helmet that held four buzzers at the head, one in each quadrant at the approximate position of a hatband. (This equipment was devised to be operated with a code and to replace voice communication between the student pilot and his instructor in the cockpit.) The equipment was described as operating tactually; however, it yielded a considerable bone-conducted auditory stimulus. One question arose from the possibility that a buzzer signal from the instructor and a voice message from a control tower might occur simultaneously and bone-conducted auditory component might serve to mask the air-conducted stimulus. As a relevant test, a recording was prepared of Forms A and B of the multiple-choice intelligibility tests. The lists were read by one voice over a carbon microphone and later mixed with noise to the extent that when played back at "test level" the intelligibility of the recording was 66%. Twenty-four experimental subjects heard the recording in the presence of the masking signal that resulted from one of the buzzers of the helmet sounding continuously.

Results. The mean intelligibility score of the "66%-record" was reduced to 52% when the recording was heard in the presence of one buzzer. When the same condition was "continued" with listeners in 110 db of simulated aircraft noise, propeller type, the mean intelligibility score of the recording was 41%. Thus, the buzzer reduced the score of the recording 14% and the added masking effect of the room noise on the intelli-

gibility score was 11%. The masking effect of the front (forehead) buzzer in this series of comparisons was not statistically different from the effect of either 110 db of room noise or the in-circuit noise of Device 8-1, a "package" that provides the major physical requirements for voice communication training (15).

APPLICATION 3: RESPONSE READINESS

The experience is common of grasping the meaning of a heard statement "more slowly than normal" some moments after the signal has been received. This experience is supposedly more frequent under adverse conditions of listening than under favorable ones. Intelligibility tests are not "speed" tests and a correct response that is made two sec. after a stimulus is heard is scored the same as a correct response that is framed simultaneously with the reception of the stimulus. The multiple-choice intelligibility tests require but a brief time for indicating a response, only sufficient for drawing a line through a word. Thus, if more time is required to recognize word stimuli under some conditions than under others, the effect might be quantified by controlling the intervals between successive stimuli. In turn, if the effect is related to some portion of the communication equipment, the obtained measurements could be criterion scores relevant to the efficiency of the equipment.

The Response Readiness test is under continuing construction as an instrument to measure the relative capacity of a piece of equipment to be "understood" rapidly. As a first approximation, the latter six of 12 recorded multiple-choice intelligibility tests were so edited that progressively shorter time intervals occurred between successive groups of words. The first six lists, of the 12, had six-sec. intervals throughout between the initiation of one test phrase and the initiation of the next. The last six lists had progressively diminishing intervals from 3.5 to 0.6 sec. from the termination of one phrase to the initiation of of the next. Three voices rotated in recording the 12 speaker lists from Form C.

A. A comparison of two microphones. The recorded Response Readiness test fed two comparison microphones--one narrow and the other broad band--that were positioned at a fixed and equal distance from an "artificial voice" (loud-speaker) located in a sound treated mixing room. The room noise was 100 db of white noise (General Radio meter, C scale). Both the recorded signal and the noise fed the two microphones yielding two independent signals going through two matched channels and alternately fed the earphone circuit of panels of listeners who sat in another room in 114 db of simulated aircraft noise, propeller type. The signal level was determined by a 1000 c.p.s. tone that was recorded at peak r.m.s. speech level. This was 104 db at the headset when fed by the narrow-band microphone and monitored by a vacuum tube voltmeter. Eleven groups of listeners of 6-10 intelligibility scores averaging "nearest 50% each" for the same one of the two microphones in the unspeeded portion of the Response Readiness test, i.e., lists 1-6.

Results. The criterion microphone under the stipulated condition above was the narrow-band with a mean intelligibility score, 50.0%; N, 40.

This score was significantly superior to the mean score (45.8%) of the broad-band microphone for the same panels (t = 2.91; 39 d.f.). However, the relationship was reversed under the circumstance of "speeded-up" speech, i.e., lists 7-12. The two means were, broad-band, 47.7%; narrow-band, 40.8% (t = 3.61; 39 d.f.). The following explanation is suggested. (a) The broad-band microphone operated under a condition of relatively attenuated "functional" signal-to-noise ratio, for it transmitted more of the total noise spectrum than did the narrow-band microphone. Thus, the 104 db signal level at the headset, as transmitted by the broad-band microphone, included more high frequency energy both within and above the speech range than the signal of the narrow-band microphone when operating at this level. (b) As the listening difficulties that were introduced by the Response Readiness test increased, the broad-band microphone operated advantageously (see Figure 1). (c) The principal result suggests that low distortion may "pay off" through facilitating more immediate recognition of words.

B. A comparison of two headsets. The Response Readiness test was employed in comparing two headsets in the manner described above relative to microphones. The signal was fed through a single broad-band microphone in the present comparison. Moreover, the comparison involved one headset with fitted ear molds and the exigencies of the situation limited the number of subjects to 13.

Like the microphones above, the two headsets could be categorized as relatively broad-band (HS-33) and narrow-band. The broad-band headset yielded scores that were not significantly different between the normal and the speeded portions of the test, +1.6% for the speeded portion. The narrow-band headset, however, was significantly penalized by the speeded portion of the test, -6.89%.

APPLICATION 4: SIGNAL-TO-NOISE RATIO

The extreme negative value of signal-to-noise ratio that has been reported in which word stimuli have been communicated is -16 db. The magnitude of this ratio, however, is dependent on the test materials that are used (12). The possibility is frequently considered that both this value and the intelligibility scores that attend various signal-to-noise ratios are functions of the transmission equipment in use, provided the language or stimulus material is of one level of difficulty. The equipment intelligibility scores for various signal-to-noise ratios might be considered as criterion scores in comparing equipments. The weight to be assigned such differences, if present, cannot be suggested a priori. One decision that affects the experimental plan at the outset is whether the side-tone aspects of the testing situation are to be included in the comparison or not. In the following illustration side-tone was not included.

A. A comparison of two microphones (results inconclusive). A high quality recording was prepared in quiet of the 12 lists of Form D of the multiple-choice intelligibility tests. One voice read the first six lists

and a second voice the remaining lists.

A comparison was made of two dissimilar microphones, one broad-band and the other narrow-band (RS-38). The two microphones were fed noise and the recorded intelligibility tests in a sound-treated mixing room. The microphones were supported so that they were one-half inch from and perpendicular to the plane of the rim at the outer flecture point of a 12-inch loudspeaker (signal source). Another loudspeaker, four feet distant, fed 100 db of white noise into the room (General Radio meter, C scale); the SPL reading at the microphone positions. The output levels of the narrow-band microphone were set such that a 1000 c.p.s. tone of peak r.m.s. voice level produced a signal-to-noise ratio of six db at the headset. The ratio was obtained when the tone and noise signals were measured separately. The corresponding signal-to-noise ratio on the wide-band microphone in this circumstance was four db. The recorded tests were presented to the microphones from the signal-source speaker and alternately through the two microphones in the mixing room to groups of 8-12 listeners per group until a total of 50 listeners had responded. The listeners were in 114 db of simulated aircraft noise, propeller type. The six successive word lists of one speaker were attenuated in two-db steps. The original level was then re-set and the process repeated for the lists of the second speaker. Subsequently, the procedure was repeated with different listeners and with simulated aircraft noise, propeller type, replacing white noise in the mixing room. Thus, six signal-to-noise ratios entered into the tests, 6, 4, 2, 0, -2, and -4 db (as measured with the narrow-band system), the voices of two speakers, and the comparisons were conducted in the presence of two types of noise in the mixing room. After the arbitrary selection of the signal-to-noise ratio at the outset, the levels were determined by altering the playback of the word signals, not the noise.

Results. Although no statistical difference was observed between the two microphones, little credence can be placed in the outcome of the test. The results are reported to illustrate a peculiarity of the multiple-choice tests. The observed means in the present instance for the six signal-to-noise ratios (microphones and noises pooled) were 25, 21.5, 22.5, 14.3, 9.8, and 5.8%. Each response on the test is a choice of four possible answers. Thus a chance score, both theoretically and in observed trials when listeners have been instructed to mark every item, is 25%. This score can either be earned or guessed. Any lower score is suspect. If the listener is "trying," a lower score may indicate that he could not follow the carrier phrase and hence made a few responses. If the listener is frustrated beyond the point of "trying," the score is meaningless. The present array of means (see Figure 2)--which incidentally approaches linearity (Lindquist, Case IV)--cannot be interpreted in terms of relative measures of equipment. These results are assumed to indicate that the present "artificial voice" was unsatisfactory in reproducing an effect comparable to in-circuit noise. The signal-to-noise ratios that were employed have operated satisfactorily with in-circuit noise.

F. A comparison of two headsets. The administration of the signal-to-noise ratio test, as described in the preceding paragraphs, was

altered only to the extent that a single microphone (RS-38) and a single noise (white) were employed in the mixing room. A group of 25 listeners apparently well motivated participated. Their listener scores made possible a comparison between two headsets, one a broad-band set (HS-33) and the other a narrow-band insert-type set, with fitted ear molds.

Results. The mean relative scores (%) yielded by the two headsets were (see Figure 3):

S/N	Broad-band	Narrow-band
+6	38.4	36.7
+4	25.9	32.2
+2	27.5	35.9
0	21.7	26.7
-2	17.2	15.6
-4	6.6	8.9

This group of scores suggests a possible way of avoiding the difficulty of "lower than guessed" scores, discussed immediately above (A). Specifically, the scores attending the broad-band headset were at least as high as "guessed" scores through the signal-to-noise ratio 2 db. The comparable and apparently superior value for the narrow-band unit accompanied 0 db. Hence, the possibility is under scrutiny that in comparisons like the present one, a statement might be made, "one piece of equipment exceeds another in a test that relies on a signal-to-noise by X db (2 db in the present instance)"--the criterion value being the level at which guessed scores are exceeded. However, tests of significance might still depend on intelligibility scores, and in the present instance the scores of 25 subjects on the overall test were not significantly different, $t = .93$.

DISCUSSION AND SUMMARY

This report serves both (a), to illustrate applications of the multiple-choice type intelligibility tests in connection with an evaluation of communication equipment and the operator and (b), to present experimental findings relevant to the equipment and the talker. For convenience, the applications are grouped under (a) voice signal, (b) the intelligibility output of equipment, (c) "response readiness," and (d) signal-to-noise ratio. Only grouping c is singularly related to the multiple-choice tests. Otherwise any advantage of the multiple-choice tests over conventional write-down tests lies in the ease and objectivity that attend scoring the printed answer forms and particularly the possibility for machine scoring.

The multiple-choice tests are differentiating among different voices and voice levels, masking level, and equipments of varying efficiency. The

tests seem to be a convenient and reliable instrument with which to test the operational efficiency of equipment.

The multiple-choice intelligibility tests have two limitations. The first is only present in greater degree in these tests than in conventional write-down tests. (a) No valid interpretation can confidently be placed on a panel score 0-25, this being below the mean guessed score. (b) The rigidity of the answer forms precludes the possibility of using the same panels of listeners endlessly. However, two sets of tests, each of which has two equivalent answer forms, reduces the seriousness of this limitation. Moreover, multiple use of the same printed form is in immediate prospect. Each set includes 24 equated tests.

The several examples that are included in this report serve as experimental results. The practice of "validating" a test and at the same time employing the results on which the test has been validated can only be employed guardedly. The several applications are presented together with the expectation that the largely consistent results add to the confidence with which the individual studies may be accepted and not viewed only as "cross validation" of a testing procedure. 1. Voice Signal Application: (a) a visual monitoring device, set to guard "normal" level, did not affect intelligibility scores significantly; (b) irregularly spaced voice signals are not less intelligible than regularly spaced ones; (c) the live voice is more intelligible than the recorded voice, but the order of relative merit is not altered by recording; (d) when talking in quiet, the speaker apparently monitors his level by airborne side-tone; and (e) decrements in the level of side-tone increase both vocal sound pressure level and intelligibility scores. 2. The Intelligibility of Equipment: (a) equipment in tower-to-plane radio communication varies in intelligibility as a function of the response characteristics of the equipment or the band width and the percent modulation; (b) the usefulness of ear wardens in improving intelligibility should be restudied; (c) a single buzzer coupled to the head to convey a tactile signal is severely masking. 3. "Response Readiness:" or speeding up the presentation rate of successive test items, the wide-band equipment appears to give a signal that is more readily understood than does narrow-band equipment. 4. Signal-to-noise Ratio: adverse signal-to-noise ratios appear to penalize the intelligibility scores of broad-band equipment.

Matters of procedure in conducting tests to yield quantitative subjective measures of intelligibility are interacting. For example, the utilization of a tape or disc recording link in the testing system is something related to whether side-tone is to be included as a contributing circumstance in the test or not. Example <u>a</u>: a microphone is engineered for an application that includes a fixed relationship between the level of side-tone and the operation of the microphone. Example <u>b</u>: a microphone is developed for a system that permits the level of side-tone to be altered. Example <u>c</u>: a microphone is developed for a system that yields no side-tone. Examples <u>a</u> and <u>b</u> probably call for maximum utilization of live voice in the testing network, at least until the vocal response of talkers as a function of side-tone variables is determined. Example <u>c</u>

can be accommodated equally well by live voice or recorded signal. However, as a guide to the laboratories and electronic centers in which microphones are developed, functions should be helpful that are derived from the administration of recorded tests that exceed the vocal dimensions that may accompany the side-tone characteristics presently engineered.

Possibly the most severe tests to which the multiple-choice intelligibility tests were subjected in the course of the trials reported here were the ones in connection with the minor alterations in the side-tone channel in quiet. First, the changes in the monitoring system were slight; and second, the scores were high, presumably in an atypical portion of the range of the test. However, the test was discriminating. The tests are being explored further to discover the relationship between score and both level and band width.

REFERENCES

1. Black, J. W., Loudness of speaking I: the effect of heard stimuli on spoken responses. The Ohio State University Research Foundation and U. S. Naval School of Aviation Medicine. Joint Project Rept. No. 2. The Bureau of Medicine and Surgery Project NM 001 053, 1949.

2. __________, The effect of room characteristics upon vocal intensity and rate. J. acoust. Soc. Amer., 21, 174-176, 1950.

3. __________, Final report in summary of work on voice communication. The Psychological Corporation, Rept. No. 21, OSRD Rept. No. 5568, 1945.

4. __________, Multiple-choice intelligibility tests. The Ohio State University Research Foundation and U. S. Naval School of Aviation Medicine. Joint Project Rept. No. 17. The Bureau of Medicine and Surgery Project NM 001 064.01.17. In press.

5. __________, Evaluation of the primary flight helmet Device 12-13 K-6 (Impact-audible signal system for flight training) Kenyon College and U. S. Naval School of Aviation Medicine. Joint Report No. 411-1-18. The Bureau of Medicine and Surgery Project NM 001 058.03.01, 1949.

6. Egan, J. P., Articulation testing methods II. Psycho-Acoustic Laboratory, OSRD Rept. No. 3802, U. S. Department of Commerce Publication Board 22848, 1944.

7. __________, Articulation testing methods. Laryngoscope, 58, 955-991, 1948.

8. Fletcher, H. and J. C. Steinberg, Articulation testing methods, Bell Syst. tech. J., 8, 806-854, 1929.

9. Fletcher, H., G. M. Raff, and F. Parmley, Study of the effort of different amounts of side-tone in the telephone set. Western Electric Co. (Engineering department), Rept. No. 19412, Case No. 120622, 1918.

10. French, N. R. and J. C. Steinberg, Factors governing the intelligibility of speech sounds. J. acoust. Soc. Amer., 19, 90-119, 1947.

11. Haagen, C. Hess, Intelligibility measurement: twenty-four word multiple-choice tests. The Psychological Corporation, Report No. 20, OSRD Rept. No. 5567, 1945.

12. Kryter, K. D., The effects of noise on man, J. Speech and Hearing Disorders, Monogr. Suppl. 1, 1950.

13. Lightfoot, C. and S. N. Morrill, Loudness of speaking: the effect of the intensity of the side-tone upon the intensity of the speaker. Kenyon College and U. S. Naval School of Aviation Medicine. Joint Project Rept. No. 4. The Bureau of Medicine and Surgery Project No. NM 001 053, 1949.

14. Miller, G. A., G. A. Heise, and W. Lichten, The intelligibility of speech as a function of the context of the test materials, *J. exper. Psychol.*, 41, 5, 1951.

15. Wearing, C. E. (Ed.), Combat Instrumentation II, OSRD and NDRC Division 17 (Transmission and reception of sounds under combat conditions) Summary tech Rept., 3, 1946.

APPENDIX A

Directions for Speakers in the Multiple-Choice Tests

This test will prove interesting to you. It will demonstrate to each of you, and to the remainder of the group who are to serve as your listeners, how efficient you are as a speaker under conditions of (describe conditions, i.e., 110-114 db of recorded aircraft noise, classroom quiet, etc.). The words you speak are to be read from the small card (5" x 7") that is on your desk. As an illustration, suppose your card contained the following words, "I am Speaker One. I say again, I am Speaker One. My name is Doe, J. G. Number 1 mortar shut assist. Number 2 blimp injure knob. Number 3 gliding battle ignite." You would read it as I did. You may have noticed that I said the identifying information, speaker number and name, without attention to the clock (sweep-hand timer). Then I waited until the hand pointed to 12 to read "Number 1 mortar shut assist." The hand goes around in 6 seconds. When it hit 12 again, I said, "Number 2, etc." Then the next time it passed 12, "Number 3, etc." This pacing of the lines allows the listeners time to respond. You may have noticed also that I read "Number 1 mortar shut assist" as a unit, a phrase, as if the words were a sentence and made sense. They usually don't, but read them as though they do, as a group, all on one breath. This permits you to read each line with the natural speech patterns of ordinary sentences. Remember: (a) read from the card; (b) say the introductory material when it is your turn to talk; then watch the clock and when the hand reaches 12 start with "Number 1, etc." Wait until the clock is at 12 again; then "Number 2, etc."; (c) read "naturally" with each group of three words-- and the word number and a digit that come before the three words-- said as a single phrase as if the line made sense, all on one breath. (Questions are called for; if none, an informal summary of the instructions is repeated.)

APPENDIX B

Directions for Listeners in the Multiple-Choice Tests

You are going to hear a series of groups of three words. From your responses we can measure both the intelligibility of the individual(s) who read the words and your efficiency as a listener in a communication situation (the particular listening environment is then described). Let's look at the front cover of your answer form. You will hear:

> "This is speaker test Form ___. I am Speaker 1; I say again, I am Speaker 1. My name is Doe, J. C. (pause) number 1 mortar shut assist (pause) number 2 blimp injure knob (pause) number 3 gliding battle ignite."

You will notice that for each word I read there are four possible choices on the cover of your answer form. You heard me say, "Number 1 mortar shut assist:" The first word after "Number 1" was mortar and appeared in the first group of four words. The second word shut was found in the second group of four words of Number 1 and the third word assist you found in the third group of four words of Number 1. Your response was to draw a line through the word you heard, making one mark in each group of four words. Erasures are permitted. (Repeat the explanation for examples Numbers 2 and 3. Questions are called for.) Each speaker will read nine (or eight in Forms A and B) groups of three words. Write the speaker's name as you hear it in the space above the list he reads. Remember, draw a line through the words you hear, or think you hear.

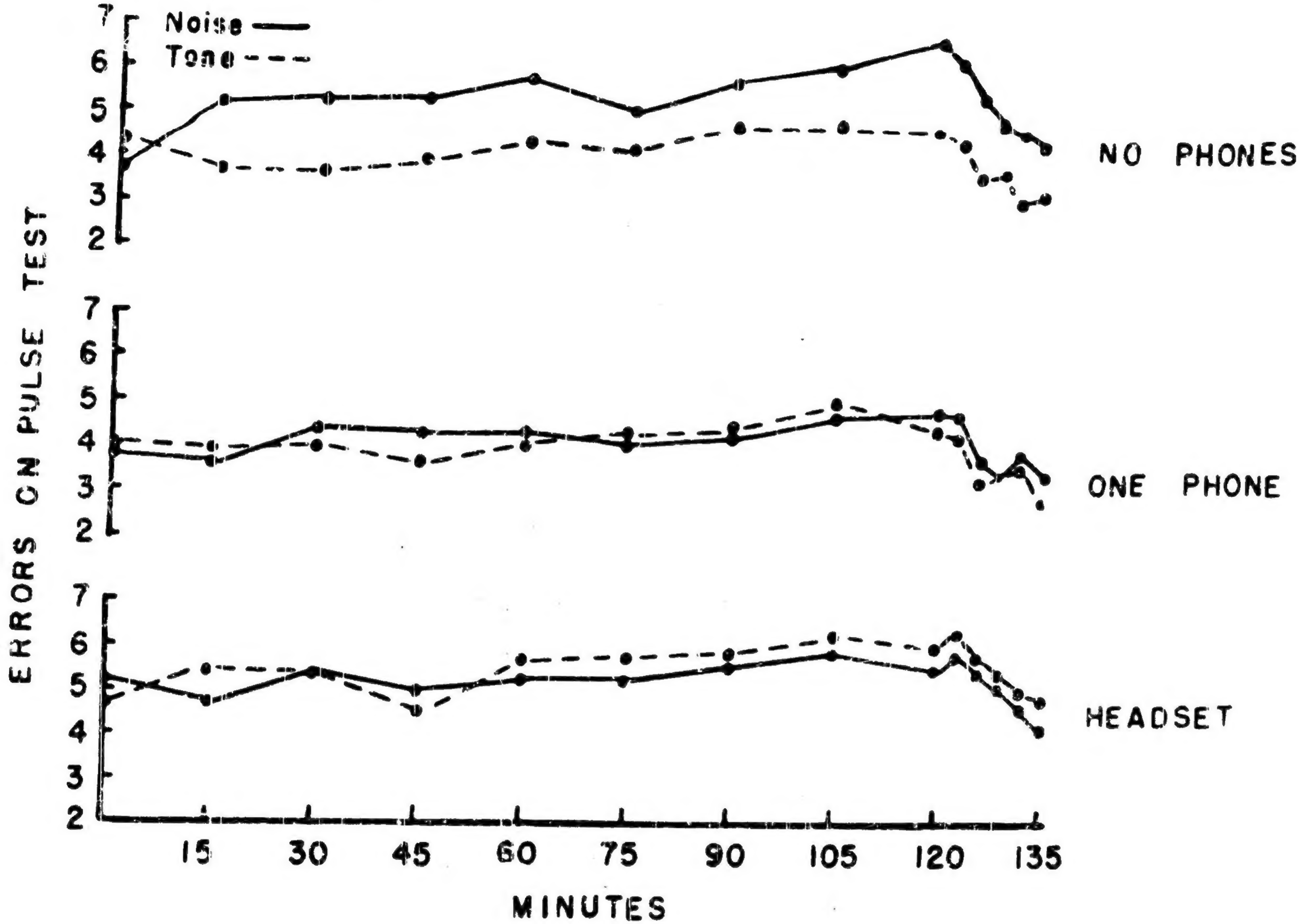

FIGURE I. MEAN NUMBER OF ERRORS ON THE PULSE TEST (NOISE AND TONE) AT SUCCESSIVE PERIODS DURING AND AFTER EXPOSURE TO NOISE.

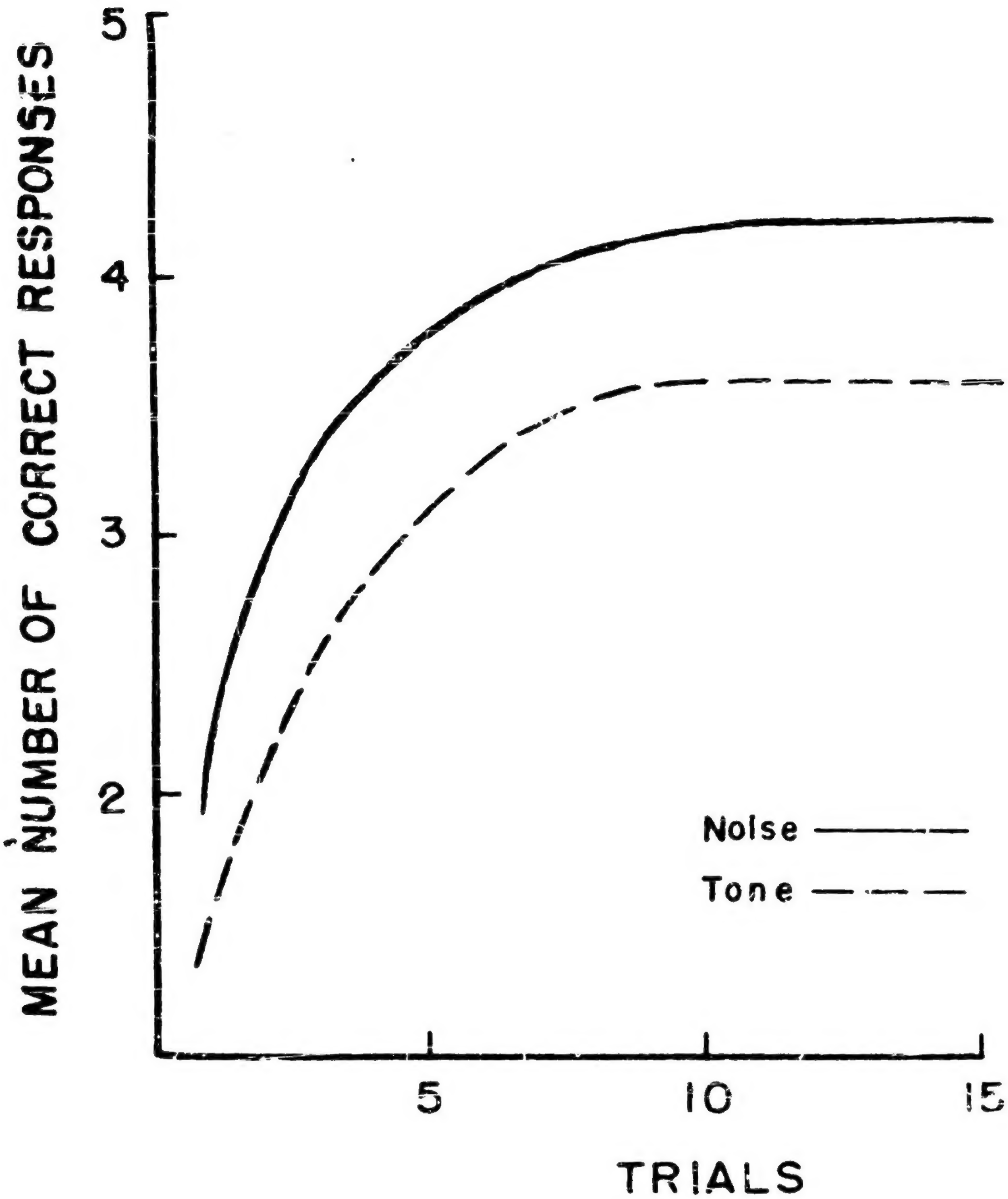

FIGURE 2. NUMBER OF CORRECT RESPONSES OF POSSIBLE 10, IN 15 SUCCESSIVE TRIALS OF THE PULSE TEST. N, SUBJECTS 120-132.

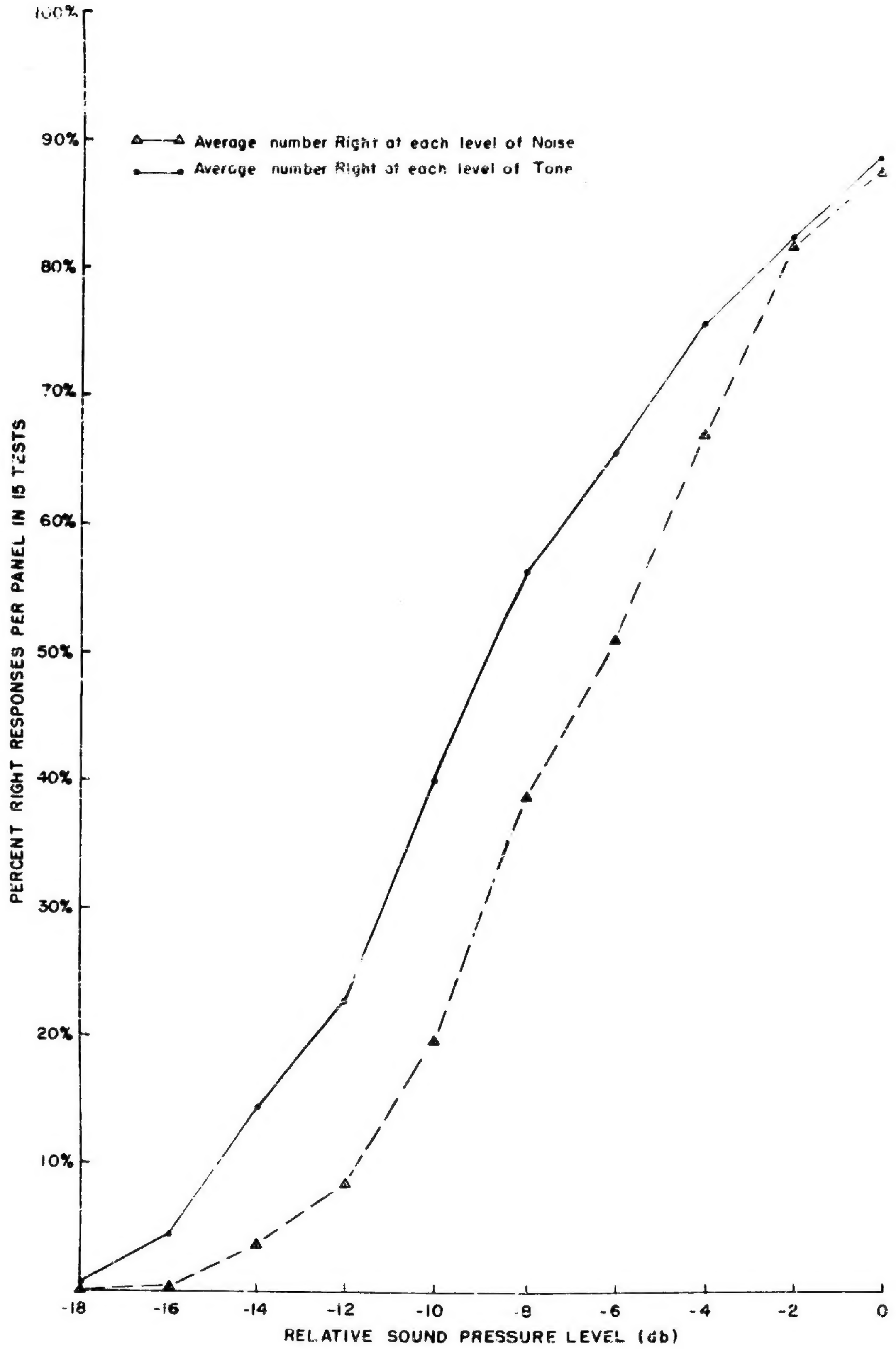

FIGURE 3. PERCENTAGES OF CORRECT RESPONSES TO ITEMS THAT REPRESENT 10 LEVELS OF SOUND PRESSURE. N, SUBJECTS, 120-132; PRESENTATIONS TO EACH SUBJECT, 15.